MS# BJP1124-R2

Running Head: INHIBITION AND INTERFERENCE

IN PRESS: BRITISH JOURNAL OF PSYCHOLOGY (March, 2012)

The role of outcome inhibition in interference between outcomes: A contingency-learning analogue of retrieval-induced forgetting

Miguel A. Vadillo[1], Cristina Orgaz[1], David Luque[2],

Pedro L. Cobos[2], Francisco J. López[2], & Helena Matute[1]

[1]*Universidad de Deusto*
Bilbao, Spain

[2]*Universidad de Málaga*
Málaga, Spain

Mailing address:
Miguel A. Vadillo
Facultad de Psicología y Educación
Universidad de Deusto
Apartado 1, 48080 Bilbao, Spain
Fax: + 34 – 94 – 4139089
e-mail: mvadillo@deusto.es

Author Note. Support for this research was provided by Grants PSI2011-26965 and PSI2011-24663 from Ministerio de Ciencia e Innovación (MICINN) of the Spanish Government, Grant IT363-10 from the Basque Government, and Grant P08-SEJ-3586 from Junta de Andalucía. We thank Fernando Blanco, Joaquín Morís, Nerea Ortega, and Ion Yarritu for their insightful comments on this experiment. Correspondence concerning this article should be addressed to Miguel A. Vadillo, Departamento de Fundamentos y Métodos de la Psicología, Universidad de Deusto, Apartado 1, 48080 Bilbao, Spain. E-mail: mvadillo@deusto.es

Abstract

Current associative theories of contingency learning assume that inhibitory learning plays a part in the interference between outcomes. However, it is unclear whether this inhibitory learning results in the inhibition of the outcome representation or whether it simply counteracts previous excitatory learning so that the outcome representation is neither activated nor inhibited. Additionally, these models tend to conceptualise inhibition as a relatively transient and cue-dependent state. However, research on retrieval-induced forgetting suggests that the inhibition of representations is a real process that can be relatively independent of the retrieval cue used to access the inhibited information. Consistent with this alternative view, we found that interference between outcomes reduces the retrievability of the target outcome even when the outcome is associated with a novel (non-inhibitory) cue. This result has important theoretical implications for associative models of interference and shows that the empirical facts and theories developed in studies of retrieval-induced forgetting might be relevant in contingency learning and vice versa.

Inhibition is a key concept in many areas of learning and memory research. To process information efficiently, humans (and other animals) need to know both when to expect a given event and when to expect its absence. Not surprisingly, most accounts of associative learning have dealt with the problem of inhibition and have included inhibitory mechanisms in their theoretical assumptions (e.g., Bouton, 1993; Dickinson & Burke, 1996; Hull, 1943; Konorski, 1948; Mackintosh, 1975; Pavlov, 1927; Rescorla & Wagner, 1972; Van Hamme & Wasserman, 1994, Wagner, 1981).[1]

Inhibitory processes are particularly relevant to understanding the cognitive mechanisms involved in situations in which a given cue is associated with different outcomes at different moments. Extinction, latent inhibition, and counterconditioning are some well-known examples of the learning phenomena that are studied within this paradigm (see, for example, Kerkhof, Vansteenwegen, Baeyens, & Hermans, 2011; Pineño, De la Casa, Lubow, & Miller, 2006; Rosas & Callejas-Aguilera, 2006). Although there are some important differences among these effects, from a general point of view, all of them involve situations in which a cue, A, is paired with one outcome, O1, in Phase 1 and with a different outcome, O2, in Phase 2. The usual result is that recall of one of the associations is reduced as a result of the training of the other association.[2] Given the remarkable similarities among all these effects, we will use the general name of *interference between outcomes* to refer to them collectively.

To cope with these sorts of situations, organisms must be able to process inconsistent information. Traditional theories of associative learning (e.g., Rescorla & Wagner, 1972) have assumed that such situations are resolved by deleting the knowledge of the A-O1 association as the A-O2 training proceeds. However, in light of the abundant literature showing that these initial associations survive after interfering training (see, for example, Matute, Lipp, Vadillo, & Humphreys, 2011; Matute, Vegas, & De Marez, 2002; Rosas,

Vila, Lugo, & López, 2001; Vadillo, Vegas, & Matute, 2004), current theories have proposed that organisms solve these conflicting associative structures by means of inhibitory associations. For example, Bouton's (1993, 1997) retrieval model assumes that after the A-O1 association has been learned, learning the novel A-O2 association requires both the formation of a new excitatory link between A and O2 and the creation of an inhibitory association between A and O1 that counteracts the previously learned A-O1 excitatory association (see Figure 1).

Although Bouton's model (1993, 1997) and some of the previously mentioned theories (e.g., Dickinson & Burke, 1996; Wagner, 1981) do stress the role of inhibitory associations in their explanations of interference between outcomes, it is unclear whether these models assume that O1 enters into an inhibitory state as a result of the action of these inhibitory links. From their point of view, the only purpose of the inhibitory A-O1 link is to suppress the action of the previously learned excitatory A-O1 association without actually deleting it. In other words, according to these accounts, the excitatory and inhibitory links cancel each other, so the presentation of A neither activates nor inhibits the representation of O1. Thus, interference between the outcomes involves some kind of inhibitory learning but not necessarily real outcome inhibition.

Recent studies of retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994; Anderson & Spellman, 1995) provide an interesting insight into the inhibitory processes that are likely to intervene in the interference between outcomes. In a typical retrieval-induced forgetting experiment, the participants are first exposed to a series of category-exemplar pairings (e.g., Fruit-Banana, Fruit-Orange, Vehicle-Car, and Vehicle-Plane). During the second phase, they are asked to actively retrieve information about some of these exemplars, with the category name and the first letters of the exemplar name as retrieval cues (e.g., they are asked to fill in the gap in Fruit-Ba___). Finally, during the test

phase, the participants are asked to retrieve all the exemplars presented in the initial study phase (e.g., Fruit-Ba___, Fruit-Or___, Vehicle-Ca__, and Vehicle-Pl___). From an experimental point of view, all of these items can be divided in three broad groups: exemplars that have been actively retrieved in Phase 2 (e.g., Fruit-Banana); exemplars that have not been retrieved during Phase 2 but that belong to a practiced category (e.g., Fruit-Orange); and exemplars that have not been retrieved during Phase 2 and do not belong to a practiced category (e.g., Vehicle-Car and Vehicle-Plane). These three categories of items are named Rp+, Rp-, and Nrp, respectively. As might be expected, Rp+ exemplars are usually recalled better than Nrp exemplars; that is, the active retrieval of Rp+ items during Phase 2 enhances their retrievability at testing. However, the most noteworthy result is that the recall of the Rp- exemplars is impaired relative to the Nrp exemplars; that is, training for the Rp+ exemplars results in both improved recall of these items and some inhibition of the related but untrained Rp- exemplars. This retrieval-induced forgetting effect has been observed in a wide variety of settings using very different materials, such as shapes and colours (Ciranni & Shimamura, 1999), sentences (Gómez-Ariza, Lechuga, Pelegrina, & Bajo, 2005), personality traits (Macrae & MacLeod, 1999; MacLeod & Macrae, 2001), eyewitness-memory scenes (Shaw, Bjork, & Handal, 1995), and affective stimuli (Amir, Brigidi, Coles, & Foa, 2001).

Typical retrieval-induced forgetting experiments usually include a relatively long (about 5–20 min) retention interval between the active retrieval phase and the test phase. This feature suggests that retrieval-induced forgetting involves inhibiting representations and that this inhibition may be a long lasting effect. Additionally, extant evidence shows that retrieval-induced forgetting is a cue-independent process: once a representation has been inhibited, it remains inaccessible even if alternative retrieval cues are used to access that representation (Anderson & Spellman, 1995). In the above experiment, for example,

once the orange exemplar has been inhibited, its retrievability would remain low both when tested with the retrieval cue Fruit-Or___ and when tested with alternative retrieval cues, such as Citrus-Or___. Moreover, this type of inhibition can be observed even with testing procedures that do not depend on the presentation of retrieval cues at all, such as recognition tests (i.e., asking participants to tell whether or not an item has been presented during the training phase and measuring their accuracy and reaction time; see Hicks & Starns, 2004; Veling & van Knippenberg, 2004).

The purpose of this study was to test whether a similar inhibitory process occurs in the interference between outcomes in a contingency-learning task. The design summary of the experiment is shown in Table 1. As can be seen, the participants were exposed to A-O1 pairings in Phase 1 and to A-O2 pairings in Phase 2. The control condition consisted of B-O3 trials in Phase 1 and C-O4 trials in Phase 2 and should not have given rise to any form of interference between the outcomes. This design is standard in interference-between-outcomes experiments (see, for example, Rosas et al., 2001). Then, after a five-minute retention interval, participants were asked to learn two new associations: one between a novel cue, D, and O1, and another between an equally novel cue, E, and O3. If O1 had become relatively permanently inhibited in Phase 2, it would follow that learning the D-O1 association would have proceeded more slowly than learning the E-O3 association. Moreover, given that O1 was predicted by a novel cue, D, the delay in the development of the D-O1 association was a measure of the extent to which the O1 inhibition was a cue-independent state, similar to the ones that have been observed in retrieval-induced forgetting studies.

Notice that in our design, the potential inhibition of O1 was tested by making participants learn a new association between a novel cue, D, and O1. We needed this procedure to test whether the effect can occur in response to a novel cue. However, this

design imposed demanding conditions for detecting any inhibitions that may have been produced during the interference treatment. First, any observation of O1 inhibition depended critically on comparing the responses to D and the responses to E. Our experiences with this task suggested that participants typically gave few or no responses to novel cues. If participants did not respond on the first E trial, determining an even lower level of response in the first D trial became impossible. Thus, no systematic responses to any of the outcomes were expected to occur on the first D and E trials. In other words, any inhibition that may have occurred could only be detected from the second test trial onward. Additionally, the first D-O1 trial required some active processing of O1, which presumably at least partially released O1 from its inhibited state, thereby reducing the chances of observing differences between the conditions in the subsequent trials. Therefore, any tested differences between the D-O1 association learning and the E-O3 association learning were considered strong evidence for the intervention of an inhibitory process.

Despite these hindrances, our task had the advantage of using a typical learning measure in human contingency-learning experiments instead of the more standard recall or recognition measures that have been used in retrieval-induced forgetting studies. This measure both provided convergent evidence for the pervasiveness of the inhibitory processes explored by memory researchers and helped to bridge the gap between these two areas of research (contingency learning and retrieval-induced forgetting), which explore similar phenomena but which have remained relatively unconnected.

Methods

Participants and apparatus

One hundred and sixty-eight psychology students from the University of Deusto and the University of Málaga volunteered to take part in the experiment. The experiment was

conducted in two large computer rooms in which the participants maintained a one-meter distance from each other. The task was programmed using Visual Basic.

Design and procedure

The design of the experiment is shown in Table 1. Participants were exposed to two training phases and one test phase. During the first phase, two different cues, A and B, were paired with two different outcomes, O1 and O3, respectively. During Phase 2, cue A was paired with a new outcome, O2, and a novel cue, C, was paired with a different outcome, O4. Each variation of the trial in both phases was presented 15 times in a fixed, pseudorandom order to all the participants. After these two phases, the participants spent five minutes filling in a clinical inventory unrelated to the present experiment. During the final testing phase, two novel cues, D and E, were paired with O1 and O3, respectively. The participants were exposed to three D-O1 trials and three E-O3 trials. To avoid any trial-order effect at test, we constructed two different pseudorandom test sequences for the D-O1 and E-O3 trials (D-E-D-D-E-E and E-D-E-E-D-D), which were counterbalanced across the range of participants.

The general procedure was similar to the one used by Luque, Morís, Cobos, and López (2009). First, the participants read the instructions and had the opportunity to ask questions. The participants could earn points by wagering on each trial. To do so, they had to learn the relationships between coloured rectangles and fictitious plants that played the roles of cues and outcomes, respectively. The colours of the rectangles used as the A-E cues were blue, yellow, pink, green, and brown (partially counterbalanced following a Latin-square design; therefore, all the colours played the role of all the cues, even though not all the potential combinations of colours and cues were used). The stimuli used for outcomes O1-O4 consisted of four pictures of plants with fictitious names underneath:

Kollin, Dobe, Yamma, and Lettsu. Their roles as the abstract outcomes shown in Table 1 were also partially counterbalanced following the same Latin-square procedure.

Before each trial, the rectangle at the middle top of the screen, in which the colour cues would later appear, was shaded in grey, indicating that the cue had not yet been presented. Below this rectangle, at the bottom of the screen, four small squares were shown, depicting the photographs of the four plants (which played the outcome role), each one with a text label indicating the response keys assigned to the individual plants (number keys 1, 2, 9, and 0). This information was shown before each trial; therefore, upon presentation of the cue, the participants needed to remember which response key was assigned to each plant. The matching of the response keys (1, 2, 9 or 0) to the plant names (Kollin, Dobe, Yamma, and Lettsu) was kept constant for all the participants.

After this sequence, the coloured rectangle that played the role of the cue in the trial appeared for 2.5 s, and the participants had the option of wagering points on each outcome plant by pressing the appropriate keys. Once the 2.5 s time had elapsed, the cue disappeared, as indicated by the rectangle returning to its grey colour. To respond, the participants placed their wagers by pressing one or more of the response options located at the bottom of the screen. When a response key was kept pressed, the points wagered for the corresponding option increased continuously, as indicated by the movement of a scroll-bar placed below the picture of each plant and by numbers in a text box increasing across a range from 0 to 100.

In each trial, the participants earned as many points as they wagered on the correct plant and lost as many points as they wagered on the incorrect plants. After each wager, the participants were told which plant was the correct one in that trial (the correct plant remained visible), the amount of points earned or lost in that trial (in a text box on the

centre of the screen) and their current total point balances (in a text box on the top of the screen). The participants were then asked to press the space bar to proceed to the next trial.

Results

Data selection and dependent variables

To guarantee that the participants had understood the instructions and paid enough attention to the task, we removed those participants who wagered zero points for the correct option from the sample in the last instance of each type of trial. Using this data selection criterion, eight participants were excluded from subsequent analysis (4.76% of the total sample).

The performance of participants in this task can be measured with different dependent variables. Perhaps the first one would be the number of correct responses given by participants in response to each cue. In the case of the training trials, this would equal to measuring the number of points correctly wagered for Outcome 1 in response to A and the number of points correctly wagered for Outcome 3 in response to B during Phase 1, as well as the number of points correctly wagered for Outcome 2 in response to A and the number of points correctly wagered for Outcome 4 in response to C during Phase 2. However, an important problem with this dependent variable is that it is not completely free from the influence of factors that can be more related to strategic gambling than to pure learning or memory processes. For instance, some participants might decide to wager for a particular outcome only when they are completely sure that their response will be reinforced, while other participants might wager for different outcomes when they are not sure about the correct response. For example, wagering 10 points for Outcome 1 does not mean the same for one participant who also wagered 10 points for the other outcomes than for a different participant who only wagered for that specific outcome. Not surprisingly, many learning experiments conducted with similar tasks rely on a slightly different dependent measure

that takes into account not only the responses to the correct outcome, but also the responses to the incorrect ones (e.g., see Rosas et al., 2001). In previous experiments conducted with this specific task (e.g., Luque & Vadillo, 2011), we have also found that the number of points earned (i.e., the number of correct responses *minus* the number of incorrect responses) can provide a more precise assessment of learning processes. Therefore, we used this measure as our main dependent variable in the present experiment. However, we also conducted the same analyses on the number of correct responses and briefly report their results.

Training phases

The upper panel of Figure 2 shows the mean number of points earned by participants through the two learning phases, collated by blocks of three trials. A 2 (Cue: A vs. B) x 5 (Block: 1-5) analysis of variance (ANOVA) conducted on data from Phase 1 yielded a significant main effect for the Block factor, $F(4, 636) = 315.07$, $p < .001$, indicating better performance as training proceeded, and an unexpected Cue x Block interaction, $F(4, 636) = 7.03$, $p < .001$. This interaction represented the differences between the responses to A and B in some of the trials. However, these differences did not systematically favour responses to one cue over the other, as shown by the absence of a main effect for the Cue factor, $F(1, 159) < 1$. Moreover, there were no significant differences between A and B in the last block, $t(159) = 0.59$, $p = .558$, showing that the responses to both cues eventually converged to the same asymptotic level of performance.

A similar 2 (Cue: A vs. C) x 5 (Block: 1-5) ANOVA conducted on the number of points earned during Phase 2 revealed significant main effects for the Cue, $F(1, 159) = 304.76$, $p < .001$, and the Block factors, $F(4, 636) = 874,17$, $p < .001$. Moreover, these effects were qualified by a significant Cue x Block interaction, $F(4, 636) = 263.2$, $p < .001$. This interaction was due to the participants' performance being less accurate for cue A

than for cue C early in Phase 2; performance became equally accurate for both cues by the end of the training. *T*-tests confirmed that the number of points earned during the first four blocks of the A trials was significantly lower than the number of points earned during the same C trials, all *t*s(159) > 2.83, *p*s < .01. However, these differences disappeared during the last block of trials, $t(159) = 1.26$, $p = .21$. This poor performance in response to A was a natural consequence of the previous pairings of cue A with a different outcome in Phase 1; it is perfectly consistent with previous demonstrations of proactive interference in memory experiments (Keppel & Underwood, 1962) and similar associative learning studies (Amundson, Escobar, & Miller, 2003; Castro, Ortega, & Matute, 2002).

The lower panel of Figure 2 shows the mean number of correct responses for the same training trials. As can be seen, the pattern of results is virtually identical. When the previous ANOVAs and planned comparisons were conducted on these data, the pattern of significant effects was exactly the same.

Test phase

The mean number of points earned across the three test trials is depicted in Figure 3. A 2 (Cue: D vs. E) x 3 (Trial: 1-3) ANOVA conducted on this measure yielded a significant effect of the Trial factor, $F(2, 318) = 372.87$, $p < .001$, but no significant Cue x Trial interaction, $F(1, 318) = 1.29$, $p = .28$. Most importantly, the main effect of the Cue factor was statistically significant, $F(1, 159) = 4.04$, $p < .05$. The significant main effect of Cue reflects that the mean number of points earned across the test trials was significantly larger for Cue E ($M = 24.18$, $SEM = 0.91$) than for Cue D ($M = 21.85$, $SEM = 1.15$), a result that supports the hypothesis that the interference-between-outcomes treatment resulted in the inhibition of the O1 representation.

Table 2 summarizes the number of points bet to each of the outcomes during test trials D1-D3 and E1-E3. In order to check whether similar results are obtained when the

number of correct responses (instead of points earned) are analysed, we conducted a 2 (Cue: D vs. E) x 3 (Trial: 1-3) ANOVA on this dependent variable. This analysis yielded a significant main effect of the Trial factor, $F(2, 318) = 575.75$, $p < .001$, and no significant Cue x Trial interaction, $F(2, 318) = 0.58$, $p = .56$. Most importantly, the main effect of the Cue factor was still marginally significant, $F(1, 159) = 2.62$, $p = .10$. Although this effect misses statistical significance by traditional standards, it is nevertheless noteworthy that it approached significance, given that this variable does not correct for indiscriminate responding and it is, therefore, less valid and trustworthy than the number of points earned.

As discussed in the Introduction, no inhibitory effects were expected in the first test trial because of the experimental design. However, we decided to further explore this issue by analysing a different dependent variable. Instead of comparing the number of points earned during the first D trials and the first E trial, we tested whether the participants avoided responding to O1 compared with the control O3 in the first trial of the test phase with the novel cues (D for some participants and E for others, depending on the counterbalancing condition of each participant). As expected, this analysis showed no significant difference between the number of O1 responses ($M = 3.26$, $SEM = 0.61$) and the number of O3 responses ($M = 3.57$, $SEM = 0.68$), $t(159) = 0.32$, $p = .75$. The low number of responses in this trial showed that the participants were reluctant to respond to novel cues, a feature that precluded any serious attempt to detect inhibition in the first test trial.

It is possible to think at least of one alternative explanation for this pattern of responding at test that is not dependent on inhibitory processes.[3] During Phases 1 and 2, participants can learn that the cue predicting O1 can also predict O2. This can interfere with learning the D-O1 association (and responding according to it) in several ways. For example, once participants see that D predicts O1 they can infer that O2 might appear on some trials, because in the past cues that predicted O1 also predicted O2 in some

occasions. This would produce the pattern of results depicted in Figure 3 without the involvement of inhibitory processes. A similar mechanism can operate in purely associative (but non-inhibitory) terms: As participants learn the D-O1 association, the presentation of D activates the memory of A (by virtue of a backward O1-A association), and A activates the representation of O2, which would then interfere with correctly responding to O1. An important prediction made by these alternative accounts is that the intrusion of O2 responses during the test phase should be responsible for the poor responding to Cue D. Moreover, this intrusion should become stronger as training proceeds. Figure 4 depicts the number of intrusive O2 responses to D, in contrast to the number of intrusive O2 and O4 responses to E. Contrary to these non-inhibitory accounts, there is no evidence that the number of O2 intrusions is higher for D than for E. Moreover, if anything, the number of intrusions tends to decrease as training proceeds. A 3 (Cue-Response Combination: D-O2, E-O2, E-O4) x 3 (Trial: 1-3) ANOVA conducted on these data yielded a significant effect of Trial, $F(2, 302) = 13.09$, $p < .001$, but no effect of Cue-Response Combination, $F(2, 302) < 1$, and no interaction between both factors, $F(4, 604) = 1.07$, $p = .368$.

Discussion

As mentioned in the Introduction, many current associative theories assume that organisms process outcome-interference situations using inhibitory mechanisms. In these models, however, the inhibitory associations are assumed to simply counteract the previous excitatory learning, and the representations of the irrelevant outcomes are neither activated nor inhibited. In contrast to this view, the results of the present experiment suggest that interference between outcomes can involve the inhibition of the outcome representation. Moreover, this outcome inhibition seems to be a state that does not vanish immediately after termination of the interference training and that is independent of the cue used to

access the representation of the outcome (i.e., once inhibited, the outcome remains inaccessible even when alternative retrieval cues are used). Although these properties are not anticipated by associative models, they are consistent with empirical research and theories on retrieval-induced forgetting (Anderson et al., 1994; Anderson & Spellman, 1995).

Our ability to observe an inhibitory effect in our experiment was noteworthy. To bridge the gap between standard retrieval-induced forgetting studies and contingency-learning studies, we were obliged to use a procedure and design that might have been suboptimal for observing a strong inhibitory effect. As explained above, there was no reason to expect differences in the first test trial, given that the participants could not be expected to know which outcome would be associated with each of the novel cues. Additionally, each test trial required the presentation of the outcomes, thus obliging the participants to actively process the outcomes and reducing any inhibition of their representation as the testing proceeded. Note, however, that this limitation was not a peculiarity of our experimental design but rather is common to all the testing procedures for retrieval-induced forgetting that require the explicit presentation of the inhibited item as part of the probe trial (such as recognition tests; see Hicks & Starns, 2004; Veling & van Knippenberg, 2004). Our observation of a significant difference in spite of this feature should be taken as strong evidence in favour of the intervention of inhibitory processes.

Although we have interpreted the results of our experiment in associative terms, our design and procedure admittedly did not allow us to ensure that the inhibitory effect we detected was in fact a product of associative processes. For example, the inhibition of O1 may have resulted from the participants' inferring and using rules or propositions that described the contingencies presented to them during the learning phases (Lovibond, 2004). Interestingly, the vast majority of the literature on retrieval-induced forgetting and

similar memory effects suggests that controlled, non-associative processes are likely to be involved in these inhibitory effects. For example, retrieval-induced forgetting tends to be absent or reduced in populations with executive control deficits or few working-memory resources (Aslan & Bäuml, 2010, 2011; Soriano, Jiménez, Román, & Bajo, 2009). Similarly, retrieval-induced forgetting is also reduced when participants are asked to perform a cognitively demanding secondary task (Román, Soriano, Gómez-Ariza, & Bajo, 2009). These results suggest the possibility that other features of controlled processes (Moors & De Houwer, 2006) may also operate on retrieval-induced forgetting. For example, are the inhibitory effects sensitive to the previous abstract knowledge of the participants? Do they depend on the conscious goal of inhibiting representations? In our opinion, these subtle distinctions have received more attention in contingency-learning research than in the study of retrieval-induced forgetting (see Mitchell, De Houwer, & Lovibond, 2009; Shanks, 2010). Therefore, future research on inhibition in memory retrieval may benefit from the methodological developments and theoretical debates that have occurred in contingency-learning research over the past few decades. The experimental preparation and design that we have presented here could easily be used to bridge that gap (see, for example, Cobos, López, & Luque, 2007; Luque, Cobos, & López, 2008).

Conversely, research on contingency learning can also benefit from the vast quantity of ideas and experimental data provided by the abundant literature on retrieval-induced forgetting. These studies both contain information about the general properties of inhibition and propose interesting theoretical interpretations of those effects that could potentially be applied to explain interference in contingency learning. For example, Anderson et al. (1994) have suggested that retrieval-induced forgetting can be understood either in terms of lateral inhibition (i.e., there are inhibitory associations between different

exemplars, and retrieving one of them automatically inhibits related exemplars) or in terms of active and voluntary suppression by an executive mechanism responsible for both the activation of relevant information and the inhibition of irrelevant information (see Anderson & Green, 2001). These two alternative explanations are represented in Figure 4A and 4B, respectively. It is interesting to note the remarkable differences between these two accounts and the standard explanation of current associative theories (as depicted in Figure 1). To our knowledge, none of the existing evidence traditionally invoked in support of associative theories can be used to discard the structures depicted in Figure 4A and 4B. Moreover, these kinds of alternative mechanisms have rarely been considered as potential explanations for interference phenomena in contingency-learning studies (but see O'Boyle & Bouton, 1996, for an interesting exception). Therefore, these two models taken from the literature on retrieval-induced forgetting should be carefully considered in future research as potential explanations for learning phenomena related to interference between outcomes.

References

Amir, N., Brigidi, B., Coles, M. E., & Foa, E. B. (2001). The effect of practice on recall of emotional information in individuals with generalized social phobia. *Journal of Abnormal Psychology,* 110, 76-82.

Amundson, J. C., Escobar, M., & Miller, R. R. (2003). Proactive interference between cues trained with a common outcome in first-order Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 29,* 311-322.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063-1087.

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature, 410,* 366-369.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102,* 68-100.

Aslan, A., & Bäuml, K. -H. (2010). Retrieval-induced forgetting in young children. *Psychonomic Bulletin & Review, 17,* 704-709.

Aslan, A., & Bäuml, K.-H. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 264-269.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of pavlovian conditioning. *Psychological Bulletin, 114,* 80-99.

Bouton, M. E. (1997). Signals for whether versus when an event will occur. In M. E. Bouton & M. S. Fanselow (Eds.), *Learning, motivation, and cognition: The functional behaviorism of Robert C. Bolles* (pp. 385-409). Washington, DC: American Psychological Association.

Castro, L., Ortega, N., & Matute, H. (2002). Proactive interference in human predictive learning. *International Journal of Comparative Psychology, 15,* 55-68.

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1403-1414.

Cobos, P. L., López, F. J., & Luque, D. (2007). Interference between cues of the same outcome depends on the causal interpretation of the events. *Quarterly Journal of Experimental Psychology, 60,* 369-386.

Dickinson, A. (1980). *Contemporary animal learning theory.* Cambridge: Cambridge University Press.

Dickinson, A. & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *Quarterly Journal of Experimental Psychology, 49B,* 60-80.

Gómez-Ariza, C. J., Lechuga, M. T., Pelegrina, S., & Bajo, M. T. (2005). Retrieval-induced forgetting in recall and recognition of thematically related and unrelated sentences. *Memory & Cognition, 33,* 1431-1441.

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review, 11,* 125-130.

Hull, C. L. (1943). *Principles of behavior.* New York: Appleton-Century-Crofts.

Keppel, G., Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1,* 153-161

Kerkhof, I., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2011). Counterconditioning: An effective technique for changing conditioned preferences. *Experimental Psychology, 58,* 31-38.

Konorski, J. (1948). *Conditioned reflexes and neuron organization.* Cambridge: Cambridge University Press.

Lovibond, P. F. (2004). Cognitive processes in extinction. *Learning & Memory, 11,* 495-500.

Luque, D., Cobos, P. L., & López, F. J. (2008). Interference between cues requires a causal scenario: Favourable evidence for causal reasoning models in learning processes. *Learning and Motivation, 39,* 196–208.

Luque, D., Morís, J., Cobos, P. L., & López, F. J. (2009). Interference between cues of the same outcome in a non-causally framed scenario. *Behavioural Processes, 81,* 328-332.

Luque, D., & Vadillo, M. A. (2011). Backward versus forward blocking: Evidence for performance-based models of human contingency learning. *Psychological Reports, 109,* 1001-1016.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276-298.

MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science, 12,* 148-152.

Macrae, C. N., & MacLeod, M. D. (1999). On recollections lost: When practice makes imperfect. *Journal of Personality and Social Psychology, 77,* 463-473.

Matute, H., Lipp, O. V., Vadillo, M. A., & Humphreys, M. S. (2011). Temporal contexts: Filling the gap between episodic memory and associative learning. *Journal of Experimental psychology: General, 140,* 660-673.

Matute, H., Vegas, S., & De Marez, P. J. (2002). Flexible use of recent information in causal and predictive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 714-725.

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 51-92). San Diego, CA: Academic Press.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32,* 183-246.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132,* 297-326.

O'Boyle, E. A., & Bouton, M. E. (1996). Conditioned inhibition in a multiple-category learning task. *Quarterly Journal of Experimental Psychology, 49B,* 1-23.

Pavlov, I. P. (1927). *Conditioned reflexes.* London: Oxford University Press.

Pineño, O., De la Casa, L. G., Lubow, R. E., & Miller, R. R. (2006). Some determinants of latent inhibition in human predictive learning. *Learning and Motivation, 37,* 42-65.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appelton-Century-Crofts.

Román, P., Soriano, M. F., Gómez-Ariza, C. J., & Bajo, M. T. (2009). Retrieval-induced forgetting and executive control. *Psychological Science, 20,* 1053-1058.

Rosas, J. M., & Callejas-Aguilera, J. E. (2006). Context switch effects on acquisition and extinction in human predictive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 461-474.

Rosas, J. M., Vila, N. J., Lugo, M., & López, L. (2001). Combined effect of context change and retention interval on interference in causality judgments. *Journal of Experimental Psychology: Animal Behavior Processes, 27,* 153-164.

Shanks, D. R. (2010). Learning: From association to cognition. *Annual Review of Psychology, 61,* 273-301.

Shaw, J. S., III, Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review, 2,* 249-253.

Soriano, M. F., Jiménez, J. F., Román, P., & Bajo, M. T. (2009). Inhibitory processes in memory are impaired in schizophrenia: Evidence from retrieval induced forgetting. *British Journal of Psychology, 100,* 661-673.

Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (SOCR): A formalization of the comparator hypothesis. *Psychological Review, 114,* 759-783.

Vadillo, M. A., Vegas, S., & Matute, H. (2004). Frequency of judgment as a context-like determinant of predictive judgments. *Memory & Cognition, 32,* 1065-1075.

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation, 25,* 127-151.

Veling, H., & van Knippenberg, A. (2004). Remembering can cause inhibition: Retrieval-induced inhibition as cue independent process. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 315-318.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5-47). Hillsdale, NJ: Erlbaum.

Footnotes

[1] Note, however, that many of these theories encompass radically different views of what inhibitory learning is (for a review, see Dickinson, 1980). Some well-known models, such as the comparator hypothesis (Miller & Matzel, 1988; Stout & Miller, 2007), even assume that the inhibition-related phenomena are behavioural effects that emerge from the operation of non-inhibitory associative processes. For the sake of simplicity, we will not discuss these different views of inhibitory learning in detail.

[2] This broad definition is particularly well suited for counterconditioning, an effect in which O1 and O2 stand as independent events, usually with opposite affective values. In the case of extinction, O2 is defined as the mere absence of the outcome trained in Phase 1, O1. Similarly, in latent inhibition O1 is defined as the absence of the outcome that would be trained in Phase 2, O2.

[3] We would like to thank an anonymous reviewer for drawing our attention towards this alternative account.

Table 1

Design summary of the experiment

Phase 1	Phase 2	Retention interval	Test phase
15 A→O1	15 A→O2	5 min	3 D→O1
15 B→O3	15 C→O4		3 E→O3

Note. Letters A to E denote the different coloured rectangles playing the role of cues. O1 to O4 represent the different fictitious plant names playing the role of outcomes.

Table 2

Points bet to each outcome during the test phase

		O1		O2		O3		O4	
Cue	Trial	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*	*M*	*SEM*
D	1	4.43	0.80	0.84	0.34	1.53	0.46	1.89	0.53
	2	33.11	1.49	0.52	0.29	0.41	0.25	1.31	0.48
	3	37.66	1.23	0.24	0.22	2.79	0.80	0.14	0.14
E	1	1.96	0.50	1.00	0.34	4.71	0.86	1.51	0.48
	2	0.72	0.41	0.57	0.28	35.26	1.31	0.22	0.15
	3	0.60	0.33	0.01	0.01	39.26	0.98	0.10	0.10

Figure Captions

Figure 1. Graphical representation of the associative structure underlying interference between outcomes in most current associative theories. Initial exposure to A-O1 pairings results in the development of an excitatory association between A and O1. Later training with A-O2 pairings gives rise to the development of an A-O2 excitatory association and also to an inhibitory link between A and O1. Excitatory links are denoted by straight arrows, and inhibitory links are denoted by dotted arrows.

Figure 2. The upper panel shows the mean number of points earned during Phases 1 and 2 correct responses. The lower panel shows the mean number of correct responses during the same trials. The error bars denote the standard error of the mean.

Figure 3. Mean number of points earned during each trial of the test phase. The error bars denote the standard error of the mean.

Figure 4. Mean number of intrusive incorrect responses to D and E during the test phase. Error bars denote the standard error of the mean.

Figure 5. Graphical representation of alternative accounts of interference between outcomes, based on the extant literature on retrieval-induced forgetting. In the top panel (4A), the inhibition of O1 is the product of a lateral-inhibition mechanism. In the lower panel (4B), a central mechanism is responsible for the voluntary and controlled activation of O2 and the inhibition of O1.

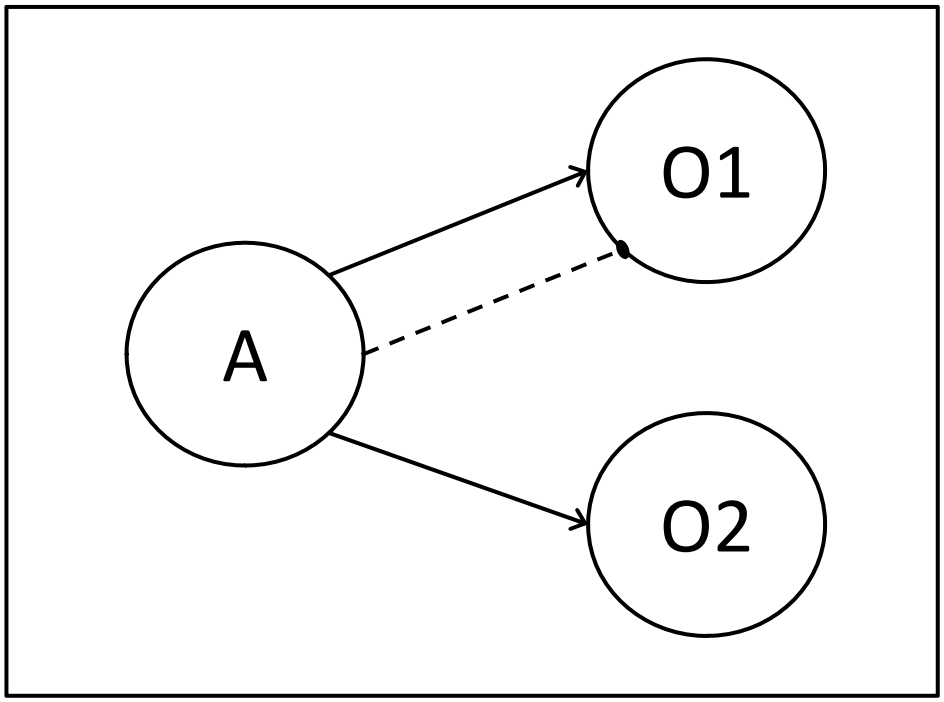
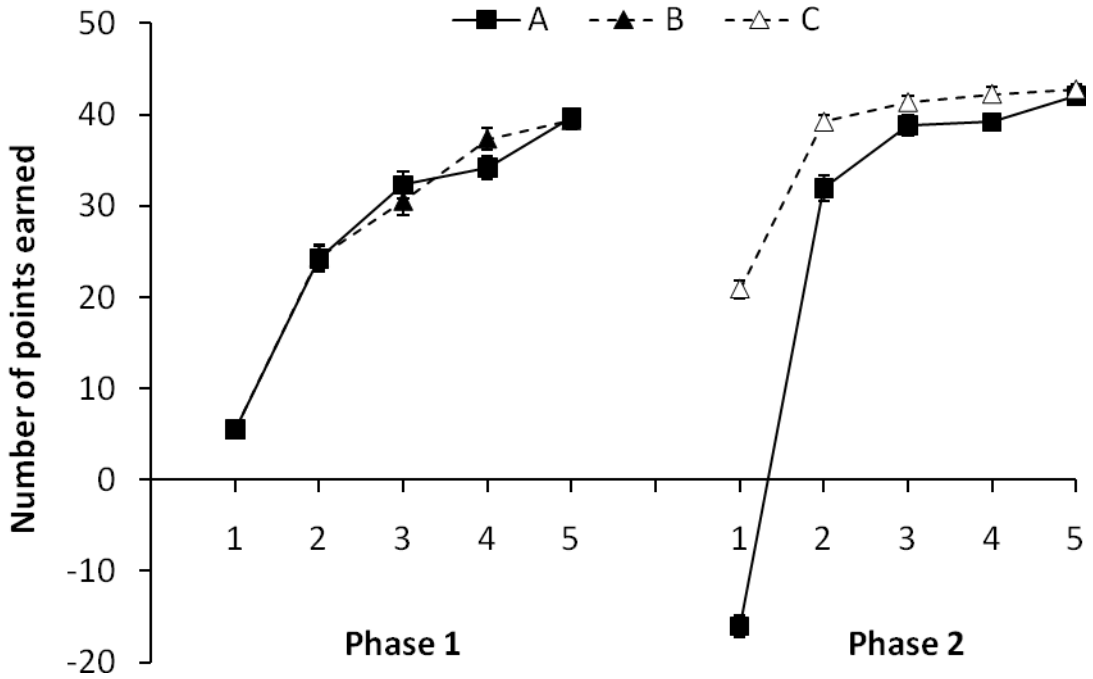


Figure #1


A
B
C
Number of points earned
50
40
30
20
10
0
-10
-20
1
2
3
4
5
Phase 1
Phase 2

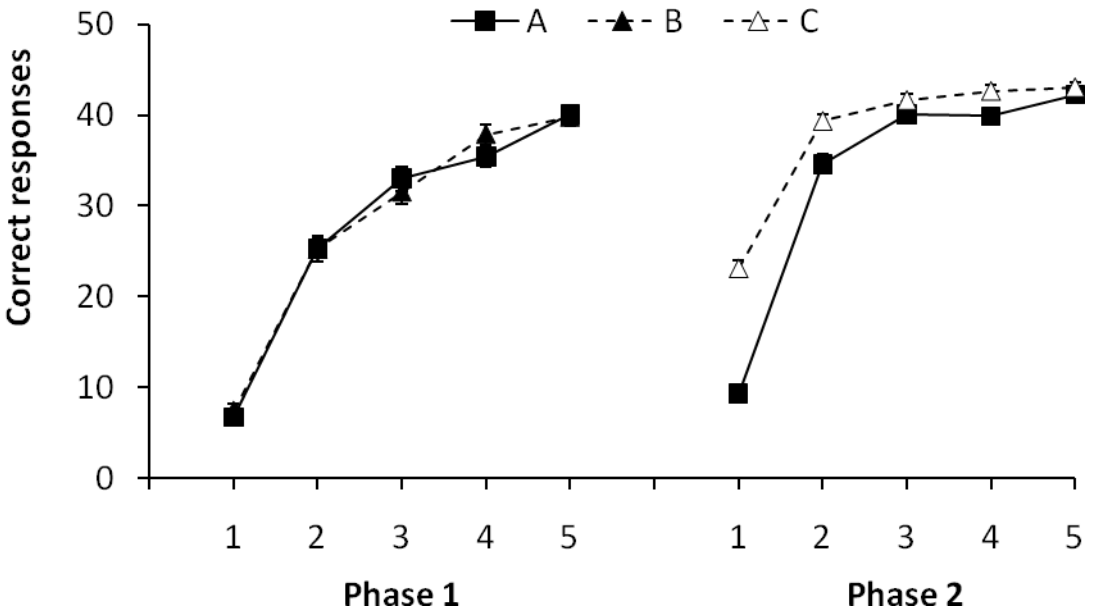

A
B
C
Correct responses
50
40
30
20
10
0
1
2
3
4
5
Phase 1
Phase 2

Figure #2

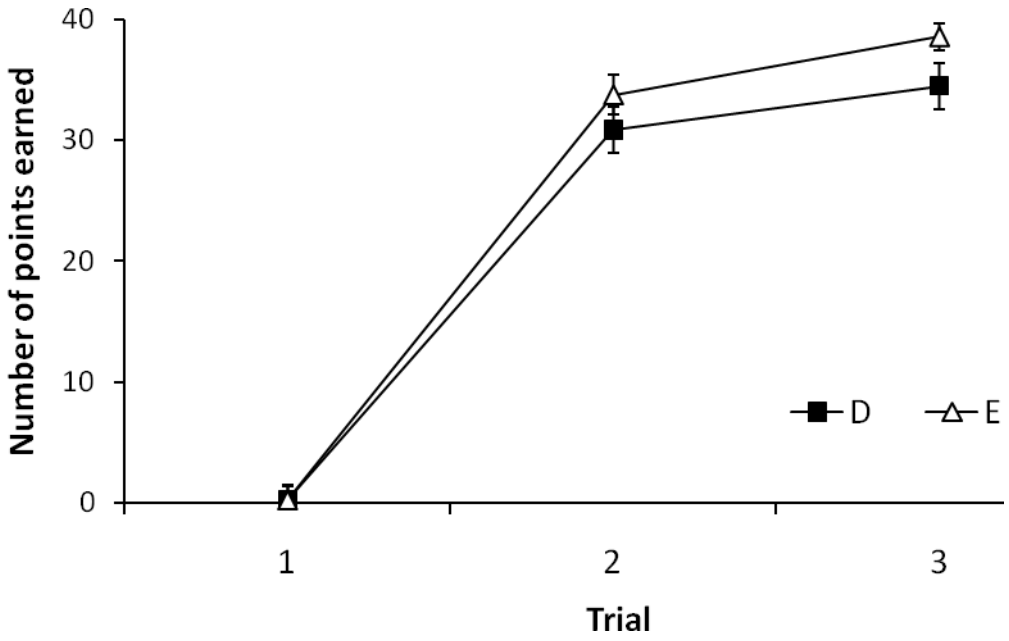


Figure #3

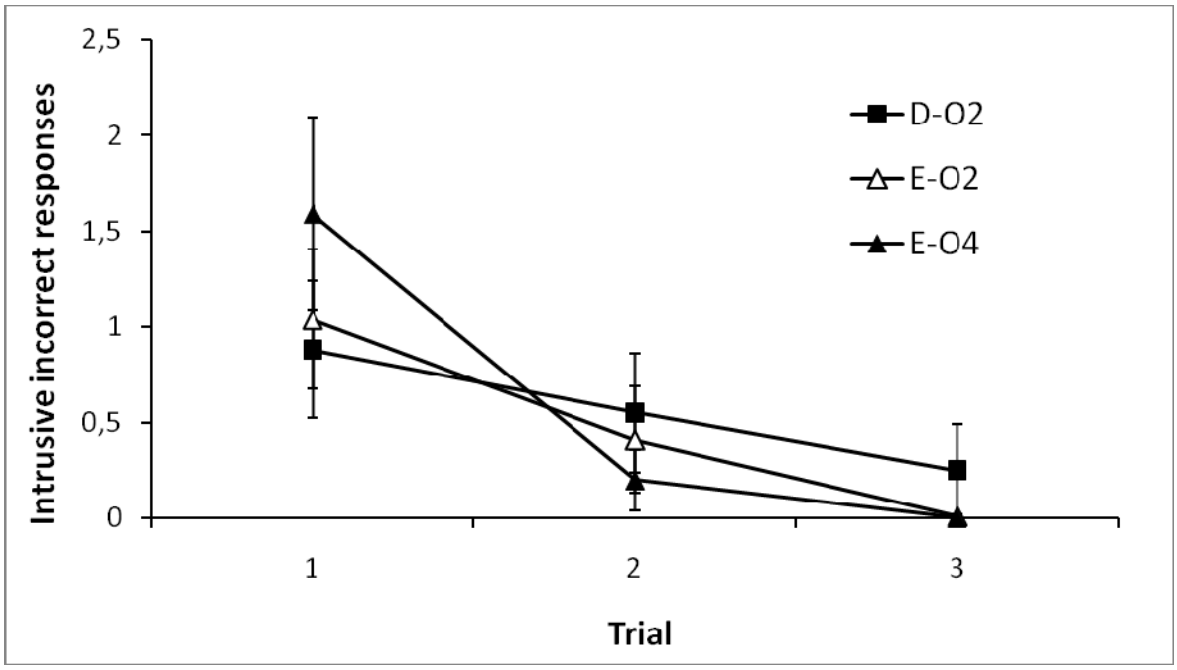

2,5
2
1,5
1
0,5
0
Intrusive incorrect responses
1
2
3
Trial
D-O2
E-O2
E-O4

Figure #4

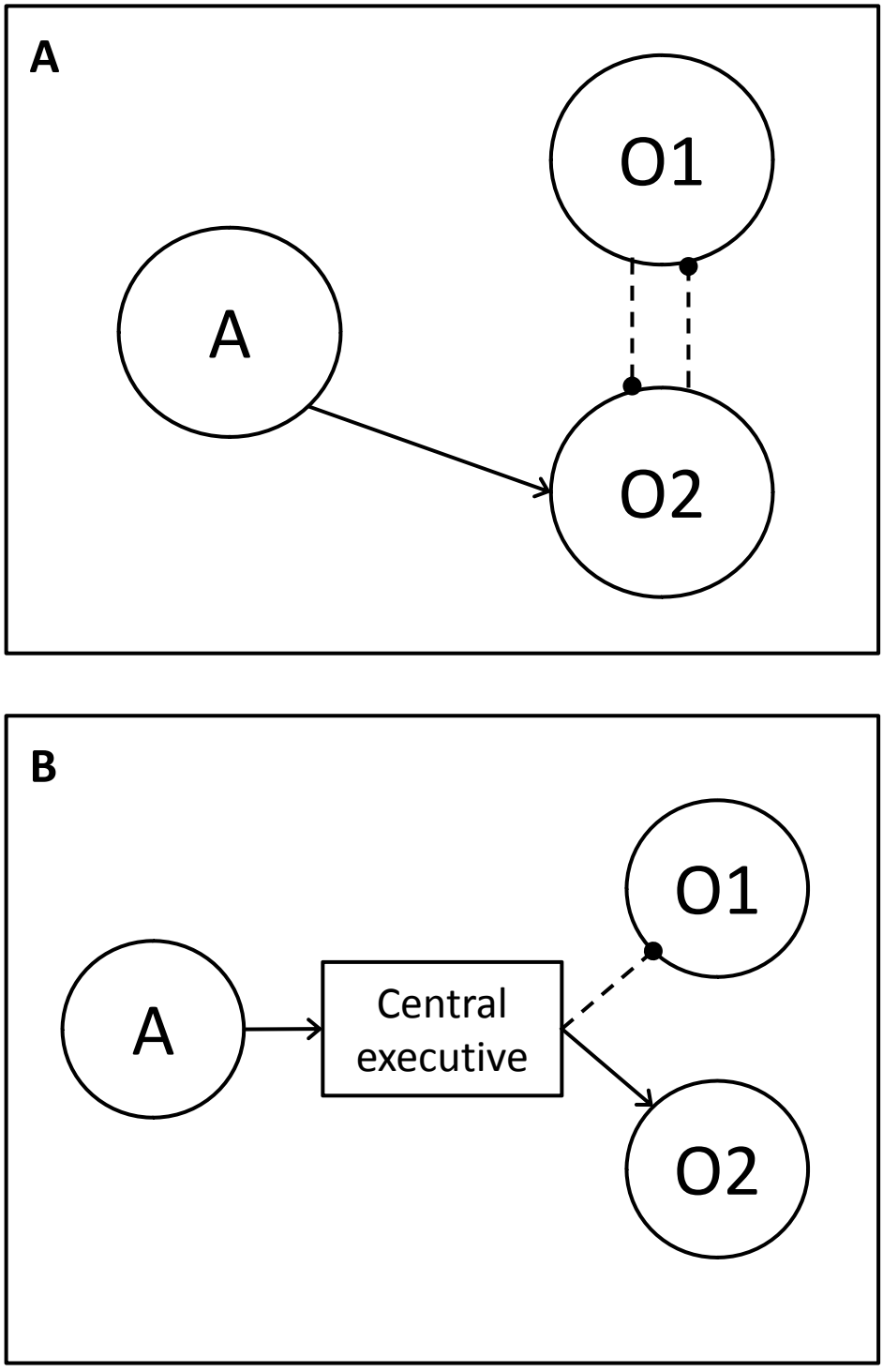


Figure #5